American Journal of Medical Genetics (Neuropsychiatric Genetics) 67:244–253 (1996)

Use of Linkage Disequilibrium Approaches to Map Genes for Bipolar Disorder in the Costa Rican Population

Michael A. Escamilla, Mitzi Spesny, Victor I. Reus, Alvaro Gallegos, Luis Meza, Julio Molina, Lodewijk A. Sandkuijl, Eduardo Fournier, Pedro E. Leon, Lauren B. Smith, and Nelson B. Freimer

Center for Neurobiology and Psychiatry and Neurogenetics Laboratory, Department of Psychiatry, University of California at San Francisco, San Francisco, California (M.A.E., V.I.R., L.B.S., N.B.F.); Fundacion de la Universidad de Costa Rica para la Investigacion (M.S., J.M.), Department of Psychiatry (A.G.), Department of History and Geography (E.F.), and Centro de Investigaciones Biologica y Molecular (P.E.L.), University of Costa Rica, San Jose; Department of Psychiatry, Calderon Guardia Hospital (L.M.), San Jose, Costa Rica; Institute of Clinical Genetics, Erasmus University, Rotterdam, Department of Human Genetics, Leiden University, Leiden, and Department of Medical Genetics, University of Groningen, Groningen, The Netherlands (L.A.S.)

Linkage disequilibrium (LD) analysis provides a powerful means for screening the genome to map the location of disease genes, such as those for bipolar disorder (BP). As described in this paper, the population of the Central Valley of Costa Rica, which is descended from a small number of founders, should be suitable for LD mapping; this assertion is supported by reconstruction of extended haplotypes shared by distantly related individuals in this population suffering low-frequency hearing loss (LFHL1), which has previously been mapped by linkage analysis. A sampling strategy is described for applying LD methods to map genes for BP, and clinical and demographic characteristics of an initially collected sample are discussed. This sample will provide a complement to a previously collected set of Costa Rican BP families which is under investigation using standard linkage analysis.

KEY WORDS: genetics, manic depression, bipolar, Costa Rica, Psychiatry

INTRODUCTION

Although twin studies and segregation analyses have strongly suggested that the etiology of bipolar disorder (BP) includes a major genetic component [Bertelsen et al., 1977; Spence et al., 1993; Pauls et al., 1992; Tsuang et al., 1994], multiple independent efforts have so far failed to definitively identify chromosomal locations for any genes responsible for this disorder [Pauls, 1993]. Attempts to identify BP genes via linkage analysis may be impeded by several possible obstacles such as etiological heterogeneity, imprecision in the definition of affected phenotypes, and uncertainty regarding mode of genetic transmission. Overcoming any of these impediments may require collection of a very large sample of families, consisting of rigorously diagnosed BP individuals drawn from genetically homogenous populations; the difficulty of accumulating such samples may be the most important limiting factor in linkage analysis of BP, as evidenced by recent efforts to develop multicenter collaborations for pedigree collection.

Genetic association studies, based on linkage disequilibrium (LD) between disease and marker loci, provide an alternative to linkage analysis for mapping disease genes, and they have two particular advantages. First, genetic parameters, which are still unknown for BP, need not be specified. Second, it is possible to collect large numbers of clearly defined cases, as samples are drawn from a population of patients rather than from families. Until recently, association studies were limited to investigation of relatively small numbers of candidate loci because it was considered impractical to use them for genome screening purposes. However, the recent development of detailed genetic maps which span the genome at close intervals now makes complete genomic screening possible, using either association studies or standard LOD score methods [Gyapay et al., 1994; Davies et al., 1994; Freimer et al., 1996]. With regard to association studies, it has recently been demonstrated that searching for chromosomal segments (defined by two or more adjacent markers) shared by affected individuals within homogenous populations provides a powerful means of mapping disease

Received for publication December 13, 1994; revision received October 6, 1995.

Address reprint requests to N.B. Freimer, Neurogenetics Laboratory, Langley Porter Psychiatric Institute, University of California at San Francisco, San Francisco, CA 94143.

genes [Houwen et al., 1994]. Houwen et al. [1994] screened the genome using this approach with currently available marker maps and identified a single region inherited identically-by-descent (IBD) by a sample of patients from The Netherlands affected with benign recurrent intrahepatic cholestasis (BRIC), a rare autosomal-recessive disorder.

LD mapping is based on the fact that, over several generations, recombination will gradually diminish the length of chromosome regions inherited IBD from a common ancestor; therefore, the degree of separation between patients that is optimal for genome screening (where markers are at least several cM apart) is much less than that which is optimal for fine-scale mapping. In most cases one will not know the exact degrees of relationship between individuals, but genealogical information can permit these to be estimated. For example, for the 3 patients used to map BRIC, the shortest distance for a few of the possible routes of relationship could be determined; based on this information, and on the assumption that unknown connections were at least one generation prior to the earliest known connection, the patients were estimated to be, on average, six generation removed from a common ancestor [Houwen et al., 1994]. If individuals used in LD screening for shared segments are much more closely related than this, it is likely that a large number of shared segments will be identified which are IBD by chance rather than due to common inheritance of the disease gene. A review of several rare inherited diseases in Finland, for which loci were originally mapped by LOD score methods, indicated that individuals separated by as many as 10–20 generations from common ancestors demonstrated conserved haplotypes of up to 10 cM around the disease gene [de la Chapelle, 1993]. The findings of BRIC and for these diseases in Finland, as well as theoretical calculations by Te Meerman et al. [1994], suggest that isolated populations in which affected individuals are about 6–20 generations removed from common ancestors are appropriate for genome screening studies using currently available sets of markers, with an average spacing of 5–10 cM between markers. As shown in this paper, the Costa Rican population is well-suited for using a search for shared segments to identify genes for BP, because the majority of the population share descent from a small number of founders, dating from the first Spanish settlement in 1569, approximately 20 generations ago.

GENETIC STUDIES IN COSTA RICA

Costa Rica is situated in Central America, between Nicaragua and Panama. The majority of the population reside in a large central valley which is geographically separated from Pacific and Atlantic coastal regions by mountain ranges (Fig. 1). This isolation caused the Central Valley population to develop independently of the populations of other Costa Rican regions, which were characterized by extensive immigration from several other countries, and are thus unsuitable for the approaches to mapping BP genes that are described in this paper.

As in other isolated populations characterized by descent from a limited number of founders, such as Finland, Costa Rica displays unusually high rates of several rate autosomal-recessive disorders [Saborio, 1992]. In particular, congenital adrenal hyperplasia has a population prevalence in Costa Rica which is several times greater than that of cystic fibrosis, usually the most common autosomal-recessive disorder. This type of "founder effect" occurs when a mutation is present in one of the original founders and is then spread by genetic drift while the population is still quite small [Lander and Botstein, 1986]. With regard to linkage disequilibrium studies, Uhrhammer et al. [1994] have demonstrated a shared haplotype on chromosome 11q22.3 in 34 of 54 apparently unrelated Costa Rican individuals with ataxia-telangiectasia, an inherited disease previously known to map to that region of chromosome 11.

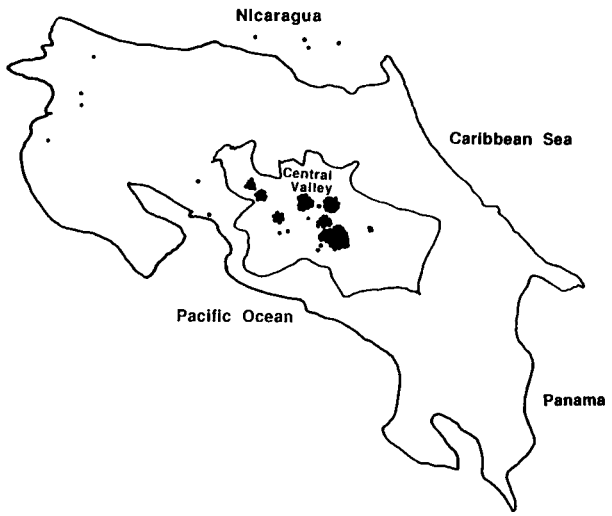

Fig. 1. Map of Costa Rica with Central Valley (home of all patients in study sample) outlined. Dots indicate place of origin of BP study subjects' great-grandparents. Four great-grandparents, who were born in Europe, are not included.

The utility of the Central Valley population for mapping disease genes was demonstrated by Leon et al. [1992], who used LOD score analysis to localize a gene for autosomal-dominant low-frequency hearing loss (LFHL1). Leon et al. [1992] carried out this analysis by using a large kindred descended from an affected ancestor in the 18th century and a smaller kindred that genealogical efforts had failed to connect with the main LFHL1 pedigree. Both kindreds were from the Central Valley. In order to show the value of the Central Valley population for linkage disequilibrium studies of complex genetic diseases such as BP, we will demonstrate the presence of linkage disequilibrium in a simple genetic illness in this region, using LFHL1 as a specific example. In this paper we suggest that most affected LFHL1 individuals in the Central Valley, including distantly related individuals and those with unknown genealogic connections, share the same extended haplotype in the LFHL1 candidate region. We will discuss how such evidence indicates that the Central Valley population is suitable for the search for shared genomic

segments in individuals who have an inherited illness in common, as previously performed for BRIC in The Netherlands [Houwen et al.,1994].

As previously mentioned, the imprecision of defining the affected phenotype is a potentially serious obstacle for linkage studies of BP. Successful gene mapping of several nonpsychiatric complex diseases, e.g., familial malignant melanoma [Cannon-Albright et al., 1992], indicate that it is prudent to focus investigation on the most severe phenotype that can be defined; for BP, this is generally considered to be BP-1 which requires at least one episode of full mania [Endicott and Spitzer, 1979]. It is likely that only a few such severely affected patients will be present in any given single family sampled for linkage analysis. As we will describe, it is possible to obtain large numbers of individuals with severe BP illness for linkage disequilibrium analysis by sampling patients from the entire population of the Central Valley. The combination of a limited number of psychiatric hospitals, universal access to treatment, maintenance of detailed medical records, and a tradition of symptom-based recording of psychiatric illnesses allows for ascertainment of virtually all BP patients in Costa Rica whose illness has been severe enough to require hospitalization.

POPULATION DEVELOPMENT IN COSTA RICA

From the initial Spanish settlement in 1569 through the late 19th century, the population of the Central Valley was separated politically and geographically from that of other regions of Costa Rica [Molina Jimenez, 1991]. The Central Valley population was mostly derived from admixture between Spanish settlers and indigenous Amerindians, occurring primarily in the 18th century. Subsequently, substantial population growth occurred in isolation, with minimal new immigration until late in the 19th century.

Details of the racial composition of the Costa Rican population were carefully documented at regular intervals from 1569–1801 (Fig. 2), as the Spanish assigned Amerindians as slaves to incoming settlers (thus requiring a detailed census of the number and location of Amerindians in 1569), and subsequently determined tax burdens based in part on the racial background of the inhabitants. This practice was abandoned after 1801, and ethnic status was not part of any subsequent government census with the exception of one in 1950.

At the time the Spanish first settled Costa Rica (ca. 1560), there were 4,800 Amerindian inhabitants in the Central Valley [Thiel, 1951]. Due to devastation by disease, and emigration from the region, the Amerindian population declined to about 1,400 in the year 1700 (224 Amerindian families were counted in the Central Valley in the census of 1697), before the advent of substantial admixture with the Spanish settlers. The majority of Costa Ricans in the Central Valley (including the Mestizos, of mixed Spanish and Amerindian ancestry) have whole or partial heritage from these original 1,400 Amerindian individuals [Acosta Vega, 1980]. The number of pure (100% Amerindian descent) Amerindians living in Costa Rica gradually declined, so that by 1950 they constituted <0.3% of the population, and lived mostly in remote areas outside of the Central Valley [Acosta Vega, 1980]. However, because of extensive intermarriage of the native American population with Spanish settlers, and to a lesser extent with Africans, in the early years of the Spanish colony, the majority of present-day Costa Ricans have partial ancestry from the original Central Valley Amerindians [Asociacion Demografica Costaricense, 1985].

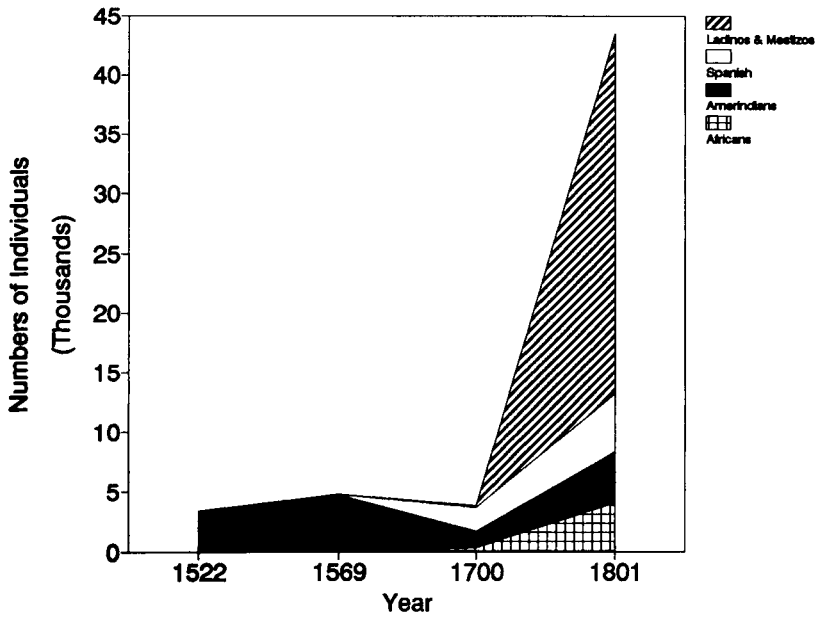

Fig. 2. Depiction of change in racial and ethnic composition of Costa Rican population during Spanish colonial period.

Genealogy of the early European (mostly Spanish) settlers of Costa Rica has been traced to country of origin from the time of the initial Spanish settlement of 1569 until the mid-19th century [Thiel, 1951; Sanabria, 1957]. Availability of extensive records of the early settlements permits this type of reconstruction of the population history of the Central Valley. The vast majority of Central Valley Costa Ricans are descended from 86 Spanish "founding families," which had a high rate of intermarriage from the start of the colony [Melendez, 1982]. From about 100–300 founding individuals, who settled the region between 1569–1575, the Spanish population grew to about 2,100 by the year 1700, and to almost 5,000 in 1801. This growth was almost entirely through reproduction, as immigration subsequent to the founding was minimal. Most of the growth in the Spanish population occurred before the Spaniards began to intermarry in substantial numbers with Amerindians or Mestizos. However, by the middle of the 18th century such matings became common, and the Mestizo population grew from about 200 in the year 1700, to about 30,000 in 1801. The latter figure is probably an overestimate, as it includes Ladinos, or individuals who may have been virtually full-blooded Spaniards but who did not have formal Spanish titles. The descendants of a small number of African slaves represent a substantial proportion of the population in coastal regions of Costa Rica. However, by the census of 1801, there were virtually no pure Africans counted in the Central Valley, and individuals with partial African ancestry comprised <10% of the population [Thiel, 1951].

Due to the minimal numerical effect of immigration on the population of Costa Rica after 1600, most current residents of the Central Valley (between 2–3 million individuals) descend from the approximately 4,000 residents of the year 1700 (including approximately 2,100 Spanish, 1,400 Amerindians, 300 individuals of African descent, and 160 Mestizos). New immigrants through the mid-19th century provided a minuscule numerical contribution to the Costa Rican population, particularly in the Central Valley [Molina Jimenez, 1991; Thiel, 1951]; e.g., out of a total population of over 105,000 in 1864, less than 1,000 individuals were immigrants. Although Costa Rica received a substantial number of immigrants from about 1890–1915, the vast majority were Afro-Caribbean individuals, excluded by law from residing in the Central Valley until 1948.

The Central Valley population is still composed mainly of native Costa Ricans. However, over the past century there has been a substantial movement of population from other regions of Costa Rica to the urban Central Valley, in particular to the capital, San Jose [Fernandez et al., 1976]; therefore, identifying individuals who are descendants of the "founding" Central Valley population requires genealogical tracing of ancestry at least to the late 19th century.

During the period in which it was relatively isolated, a high fertility rate caused the population of the Central Valley to expand dramatically, particularly during the 19th century (from about 43,000 to about 250,000). During the period from 1650–1850, the average number of descendants ranged from 5–8 children per family [Abarracin-Gonzalez, 1978]. A high birth rate has continued until the present time: in 1973 the average 40-year-old woman in Costa Rica had given birth to 7 children [Fernandez et al., 1976].

MATERIALS AND METHODS

Collection of BP Cases

The ascertainment strategy for collecting a sample of BP patients suitable for LD analysis was based on identifying individuals with well-documented diagnoses of severe BP, with typical clinical presentation and course, who were descended from the founding population of the Central Valley. This was accomplished by a two-tiered set of screening questions (Fig. 3). The first screen (screen 1) consisted of ascertainment of BP patients from a variety of inpatient and outpatient sources in the Central Valley of Costa Rica, with the most effort devoted to the National Psychiatric Hospital (NPH) and the Calderon Guardia Hospital (CGH), the only inpatient psychiatric hospitals in the country. We systematically screened medical records for all patients discharged from the NPH from 1981 through the end of 1992, and selected only pa-

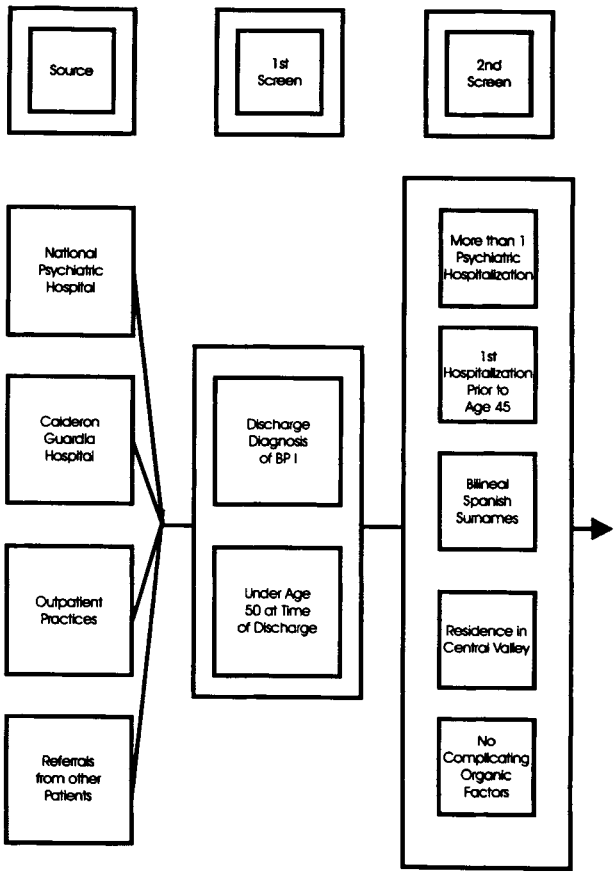

Fig. 3. Schematic representation of protocol used to ascertain and clinically evaluate BP-I patients from Costa Rican Central Valley population.

tients who had the discharge diagnosis of bipolar disorder, manic type (ICD-9 coding 296.0; records at the NPH are filed using ICD-9 coding rather than DSM-III-R coding), and who were under age 50 at the time of discharge. These criteria were designed to eliminate from further consideration all individuals except those whose diagnosis could be classified as BP-I and to decrease the sampling of individuals with a late onset of illness. The second step of the screening process (screen 2) consisted of selecting a subgroup who satisfied additional criteria. At this step we only included individuals who had at least two psychiatric hospitalizations, reasoning that those who had only been hospitalized once were more likely to have a mild or atypical form of BP. We excluded individuals whose first hospitalization was after age 45, to minimize the likelihood of identifying individuals with secondary mania, [Kramer et al., 1970], and because of evidence suggesting that early onset is as-sociated with greater familial aggregation for BP [Mendlewicz, 1976; Stone, 1989]. To decrease the number of individuals without Costa Rican ancestry, we sampled only those with a Hispanic surname for both parents (in Latin America individuals use both parents' last names). To enhance the likelihood of identifying descendants of the founding population, we limited enrollment to individuals who currently reside in the Central Valley. At this stage we also excluded individuals who had an organic complicating factor listed in their diagnosis (such as organic affective disorder and dementia).

An attempt was made to contact all patients who successfully passed both of these screening steps. Once a patient agreed to participate in the study, his or her hospitalization records were carefully reviewed by a psychiatrist who specializes in mood disorders, and detailed notes were taken, using a format that documents DSM-III-R symptoms for mania, major depressive disorder, psychosis, and substance abuse, as well as medical history and laboratory evaluations. This was done in order to verify the discharge diagnoses that were used in the first screen. If, after this detailed medical record check, a patient met clear DSM-III-R criteria for a distinct manic episode, the patient was again contacted and a blood sample obtained for DNA analysis. As a final step in collecting diagnostic data, a blinded interview was done by a bilingual psychiatrist with expertise in diagnosis of mood disorders, using a standardized semistructured interview (the Diagnostic Interview for Genetic Studies, DIGS) [Nurnberger et al., 1994]. Final diagnoses were made using a best-estimate process, which involved a consensus diagnosis by two psychiatrists, after each had independently reviewed the extracted medical records, DIGS interview, and clinical narrative for each patient.

In addition to patients ascertained through the two psychiatric hospitals, patients who met the above criteria (from screens 1 and 2) were identified through outpatient psychiatric practices and were referred to us by other study subjects. In some cases, these patients had been hospitalized before 1981 and therefore had not been selected in the screens conducted at the NPH. Patients ascertained from these sources were followed up using the same process described above for those patients located originally through hospital records.

Institutional review board approval for studies with human subjects was granted by each participating institution, and all participants signed informed consent statements.

Genealogic Assessment

In order to determine which of the subjects were descended primarily from the original founders of Costa Rica, a genealogic search was completed for each patient, using church and civic records of births, marriages, and deaths. Lineages for each patient were traced back three generations (total of 14 ancestors per patient). This allowed documentation of each subject's ancestry back to the 1880's or earlier, before there was substantial immigration and migration within Costa Rica (see earlier in Population Development in Costa Rica). Each patient was then given a weight determined by the number of Costa Rican-born ancestors out of the 8 ancestors in the great-grandparents' generation of their family. The intention of this study was to collect mainly individuals with 7 or more ancestors out of 8 in the great-grandparents' generation from Central Valley origins. Since 1892, birth registration has been centralized in San Jose, facilitating genealogic tracing to that date.

LFHL1 Haplotype Reconstruction

We evaluated the pedigree structure and genotyping results for the Costa Rican LFHL1 kindred detailed in Leon et al. [1992], and were able to identify several individuals who were separated from each other by 8–10 meioses, approximating the degree of relationship of a desirable sample for LD screening studies. Four such individuals (A–D) were randomly and blindly selected from the larger pedigree (the "FM" pedigree), in or-der to evaluate the conservation of haplotypes in the LFHL1 candidate region, between marker loci IL9 and GRL on chromosome 5. These 4 individuals can all be traced back to a single ancestor who was born in 1754, and the number of meioses separating any 2 of them from a shared ancestor range from 4–7. Leon et al. [1992] had reported an additional smaller pedigree (the "SM" pedigree) in which LFHL1 had segregated. This family was thought to be distantly related to the larger pedigree, but there were no identified connections between them, despite genealogic tracing as far back as 1754. One patient from this smaller pedigree (U-A) was randomly and blindly selected for comparison of the marker haplotype with those of the patients from the main pedigree, to provide an empirical example of haplotype conservation among affected individuals, without known relationship, who descend from the founding population of the Central Valley.

RESULTS

BP Sampling

The first screen of medical records revealed a total of 439 patients who had been discharged from the NPH with a diagnosis of bipolar disorder, manic phase, from 1981–1992 and who were under age 50 at time of discharge (records for the years 1985 and 1991 were not

available). To date, the second screen has been completed on 233 of these patients. Of these 233, 97 satisfied all of the clinical criteria of the screening process, but only 74 had known residences in the Central Valley. Six additional individuals were referred by other subjects during the course of this study; 3 of these 6 met the necessary criteria to be included in the study. Six patients from private psychiatric practices and 2 patients from the Calderon Guardia Hospital Lithium Clinic passed both screens and were included in the pool of patients to be contacted.

Patients who met the clinical and demographic criteria (N = 85) were sought in order to request their participation in the study; we were unable to locate 32 of these individuals, leaving 53 who were invited to participate in the study. Of these 53, 49 (92%) agreed to participate. After detailed review of their medical records, 48 out of the 49 participants (98%) met the DSM-III-R criteria for BP-I with at least one clear manic episode requiring hospitalization (one met criteria for BP, type 2, with only hypomanic and major depressive episodes: this patient was excluded from further study). Of the 48 final participants, 37 (77%) had received a trial of lithium during the course of their illness, 23 (48%) had received a trial of carbemazapine, and 2 persons had been treated with valproic acid. Forty-one of these patients (85%)had been on neuroleptic medication, and 16 (33%) had been on tricyclic antidepressants.

The patients studied to date include 37 ascertained from the NPH, 2 from the CGH, 6 from outpatient practices, and 3 from referrals by other participants, for a total of 48 patients.

Table I presents disease characteristics for all 48 cases who met the criteria necessary for LD mapping. Although the ascertainment process was performed independently of patients' sex, our sample yielded essentially equal numbers of men and women (M = 23, F = 25), suggesting that the prevalence of BP in Costa Rica is equivalent between the sexes. The average age at first hospitalization (M = 23.0, F = 28.9) is only slightly later than age of onset of psychiatric disorder in our patients (M = 20.5, F = 25.7). Age of onset referred to the age at which the first episode of mania or major depression occurred. Age of onset in the women appeared bimodal, with 20 patients having an age of onset between 14–29, and the other 5 having age of onset between ages 38–42. In contrast, age of onset for the men was unimodal, with all the men having an age of onset by 31, and all but 3 having their first hospitalization by age 28. Although our sampling strategy selected patients with at least two hospitalizations, most patients had been hospitalized several times (M = 6.8, F = 5.4), indicating that many of these individuals experienced substantial recurrent morbidity.

TABLE I. Clinical Characteristics of Study Sample

	Women	Men
Total	25	23
Average age	47.5	40.3
Average age of onset	25.7	20.5
Average age at first hospitalization	28.9	23.0
Average number of hospitalizations	5.4	6.8
Average length of hospital stay (days)	17.6	16.7

Preliminary genealogic studies have been partially or wholly completed for 46 individuals (Table II, Fig. 1). All 46 individuals had at least 3 Central Valley ancestors dating back to the great-grandparents' generation, and the average percentage of great-grandparents born in the Central Valley was 73%. Thirty-nine of these individuals have thus far revealed grandparents of entirely Central Valley origin. Only 5 individuals had ancestry from outside of Costa Rica, and only 2 others had ancestry from regions of Costa Rica outside the Central Valley. The vast majority of great-grandparents were born in the 19th century, some as far back as 1844.

Incidentally, the genealogic search revealed instances of mental illness in at least 2 subjects' distant ancestors (resulting in either death or divorce, and therefore on record in the church and/or civil records), an example of the detail available in Costa Rican civic records. This type of information, in addition to the centralized records of the National Psychiatric Hospital which date back to the late 19th century, provides readily available signposts with which to trace the route of inherited mental illnesses in Costa Rica.

LFHL1 Haplotype Reconstruction

Of the 4 affected individuals in the main LFHL1 pedigree selected for this study, separated from one another in the family by 8–10 meioses, 3 demonstrated a shared haplotype from marker locus IL9 to 5S119, a distance of about 28 cM (Fig. 4). The fourth individual did not share the conserved haplotype over the entire 28 cM region, but did share it from IL9 to 5S70, a distance of 7 cM. In the LFHL1 branch ("SM") without known connections, all the affected individuals shared a haplotype from FIB5–5S22 (a total of 51 cM); this region included a segment from GRL–5S119 (21 cm) that was shared by 3 out of the 4 individuals evaluated from the main pedigree. By contrast, a randomly selected chromosome from an unaffected individual (U-U) in the smaller ("SM") family shared only alleles at GRL and 5S70, two relatively uninformative markers, in common with the affected individuals studied. GRL and 5S70 are both markers with a low polymorphism content (e.g., allele 1 at marker site 5S70 has a frequency of approximately 88% in the Costa Rican population, determined by an analysis of the alleles present at that site in 60 chromosomes from 30 unaffected individuals in the LFHL1-A and LFHL1-B pedigrees), and as such will inevitably identify shared regions that are not actually haplotypes identical by descent, but rather chromosomal regions that share alleles by state. In the case of persons such as U-U, analysis of interspersed, closely-placed polymorphic markers in these individuals should reveal if they actually share a haplotype with affected individuals or not.

DISCUSSION

Linkage disequilibrium (LD) analysis as a method of initial genome screening was first proposed by Lander

TABLE II. Birthplace of Great-Grandparents of Bipolar Subjects

Central Valley of Costa Rica	Costa Rica (outside Central Valley)	Central America (outside Costa Rica)	Europe	Unknown
319	10	9	4	26

and Botstein [1986] as a substitute for LOD score methods. Although LD mapping is not a new concept, it has only recently been demonstrated to be a feasible, independent method of screening the genome to map the location of disease genes, as was shown in the case of BRIC [Houwen et al., 1994]. Since LD mapping is based on a search for chromosomal segments which are shared between affected individuals who are descended from a common ancestor, this approach requires a population in which rapid population growth has occurred without extensive immigration subsequent to its "founding" 6–20 generations in the past. The population of the Central Valley of Costa Rica, as described above, is exceptional in that over 2 million current inhabitants are descended from 224 Amerindian families and 86 Spanish families, who began to mix in the 17th and 18th centuries. Since both founding populations were small in size, it is likely that mutations introduced from either the Spanish or Amerindian gene pools would be detectable in the current Central Valley population using LD analyses. The high frequency of several autosomal-recessive disorders in Costa Rica [Saborio, 1992], the finding of a shared disease haplotype in apparently unrelated Costa Rican individuals with ataxia-telangiectasia [Uhrhammer et al., 1994], and the identification of an extended haplotype conserved between 4 out of 5 distantly related Costa Rican individuals with LFHL1 provide empirical evidence in support of our interpretation of the population history of this region. The sample of BP-I patients described above, the majority of whose ancestors are of Central Valley origin dating back to the late 19th century and who have shown no direct relationship between each other in the last four generations, offers direct evidence that an appropriate LD sample can be collected in Costa Rica.

In the case of LFHL1, if a genome screen had been performed with markers at intervals of 10 cM, using the 5 individuals evaluated above, a shared region in 4 of the individuals would invariably have been ascertained, since they share a 21-cM region. If mapping had been performed with markers spaced at 14-cM intervals, depending on the placement of the markers, a shared haplotype would have been found in a minimum of 3 out of 5 individuals (if, for instance, IL9 and 5S210 were used) and a maximum of 4 out of 5 individuals (if 5S210 and 5S119 were used). As demonstrated by Houwen et al. [1994], in evaluating patients whose common ancestry is postulated but not known, and who are separated from a common ancestor by >6 meioses, it is expected that such conserved haplotypes will be identified around disease genes but not elsewhere in the genome.

A comparison to other populations suitable for LD mapping reveals distinctive features of the Central Valley of Costa Rica's population. The individuals in whom BRIC was mapped were inhabitants of a specific region of The Netherlands whose population was known to descend for the most part from ancestors who settled that region by the 17th century, closely paralleling the situation in Costa Rica's Central Valley. As mentioned above, Finland has also displayed empirical evidence

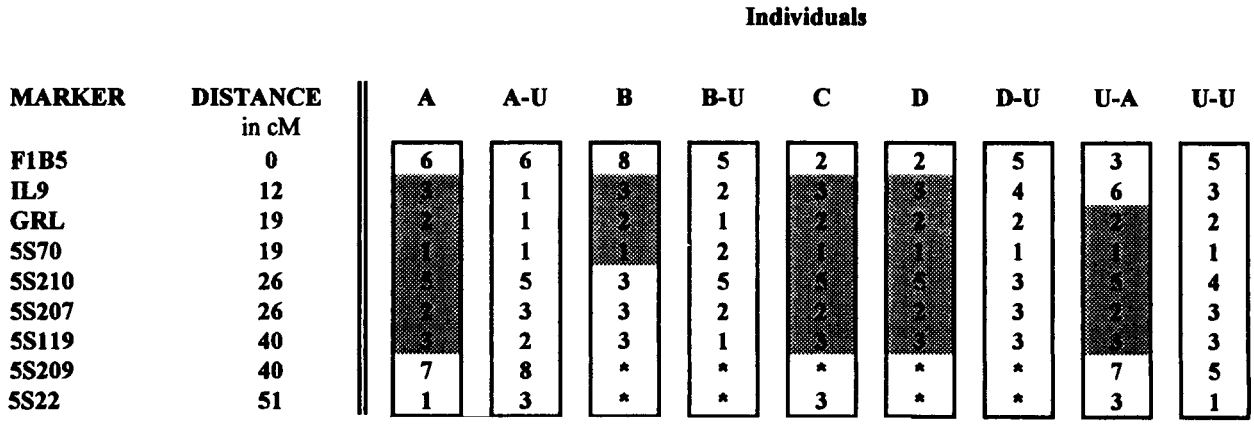

Fig. 4. Haplotypes in the vicinity of the LFHL1 locus in distantly-related members of Costa Rican pedigree in which this gene was mapped. The haplotype conserved among affected individuals is shaded. The LFHL1 gene lies between IL9–GRL. A, B, C, D and U-A are affected individuals. A-U, B-U, D-U, and U-U are unaffected individuals from the same family branches as the affected individuals (one randomly selected chromosome is displayed for unaffected individuals). Individuals U and U-A are sibs from the "SM" family, which has no documented geneological connection to the "FM" family (of which A, A-U, B, B-U, C, D, and D-U are members). Distance is measured from F1B5, using the 1994 Genethon map [Gyapay et al., 1994]. *, missing data.

for LD in inherited illnesses, due to its homogeneous founding population, large population expansion 2,000–5,000 years ago, and the relative isolation of certain rural areas [de la Chapelle, 1993]. The earlier founding of the Finnish population (80–100 generations ago) would most likely result in smaller conserved haplotypes around disease genes inherited from a "founder" than would be expected in Costa Rica, decreasing the utility of the Finnish population for genome screening by LD analysis (markers would have to be more closely spaced to pick up a shared region), but increasing its utility for fine-scale mapping once a disease region has been mapped.

Populations within French Canada (the Saguenay-Lac-Saint-Jean region of Quebec) and on the island of Reunion have also displayed LD for particular diseases [Mathieu et al., 1990; Beckman et al., 1991; Fougerousse et al., 1994]; the dates of their "founding," admixture of Amerindian and European founders, and subsequent expansion are more similar to the situation in Costa Rica, and illustrate the point that an original admixture in the founding of a population does not diminish its appropriateness for LD analysis. In fact, these three colonial populations have the advantage of more recent founding and expansion than most European and Asian communities, and therefore more closely approach the ideal situation for genome screening in a search for shared segments. For instance, when one compares Finland and Costa Rica with regard to population expansion over the last 300 years (the ideal 6–20 generations for LD mapping), the Central Valley of Costa Rica displays a much greater expansion from a smaller group of founders than does Finland (the Central Valley of Costa Rica expanded 500-fold from a starting population of 4,000 during this period, while the Finnish expanded 20-fold from a starting population of 250,000). The Saguenay-Lac-Saint-Jean region of French Canada has a more recent founding than Costa Rica (1838), and only a 6-fold expansion since 1911. Although this fits the minimal criteria necessary for linkage disequilibrium screening to succeed, the relatively small size of the population (300,000) compared to Costa Rica may limit the number of cases that have descended from any one founder [Heyer, 1994], and a genomic screen might be expected to pick up more false-positives than a similar screen in Costa Rica (where the number of generations since the founding is larger, thus allowing unlinked chromosomal regions more chances to be dispersed in the population). Reunion's founders go back at least to the 1600's, most closely approximating the situation in Costa Rica, but the relatively small population might, as in French Canada, diminish the size of the available study sample for any given inherited illness.

BRIC (the first illness to be mapped by LD analysis) and LFHL1 (our example of LD in an illness from the Central Valley of Costa Rica) are both characterized by simple Mendelian patterns of inheritance. In contrast, BP might be expected to be etiologically heterogeneous, even in a genetically isolated region, due to either allelic or locus heterogeneity or the existence of phenocopies. In fact, locus and allelic heterogeneity have already been demonstrated for at least two inherited diseases in Finland, whose population is possibly even more homogenous than that of Costa Rica [de la Chapelle, 1993; Nystrom-Lahti, 1994]. Houwen et al. [1994] demonstrated through simulation analyses that a search for shared segments would yield considerable power in detecting loci of interest even in situations of moderate etiological heterogeneity, where only 50% of the individuals shared the same disease mutation. In such cases, one seeks haplotype-sharing in an appropriate proportion of affected individuals. Hereditary nonpolyposis colon cancer (HNPCC) provides an illustration of how LD analysis might succeed in the case of a common inherited heterogeneous disease. Despite known locus heterogeneity (causative genes for HNPCC have now been cloned on chromosomes 2 and 3 [Peltomaki et al., 1993]), it was shown in a Finnish sample that a 10-cM segment on chromosome 3 was shared by affected individuals in 7 out of 13 HNPCC families, and an identical mutation was found in all 7 of these families [Nystrom-Lahti, 1994]. Closer analysis of these Finnish HNPCC families reveals that all 7 families sharing the 10-cM haplotype were descended from ancestors who had lived in a small region of south-central Finland. Of the other 6 families whose haplotypes were studied, only one was descended from this same region; 2 families from the Eastern seaboard shared a haplotype of from 1–3 cM around the causative gene, and the remaining 3 families originated in other regions and showed no haplotype sharing around this locus. In summary, approximately 50% of the affecteds shared a haplotype in a specific area of chromosome 3, and these individuals turned out to be the subgroup of HNPCC patients which had an HNPCC-related defect at that chromosomal site.

Other strategies exist for overcoming the obstacle of locus or allele heterogeneity in a population. The finding for HNPCC patients in Finland suggests that a sampling approach utilizing information on regional population history within a country can be used to preferentially select patients who are likely to inherit a disease mutation from a common ancestor. Such a strategy increases the power of LD analysis in disorders characterized by etiological heterogeneity. In Costa Rica, the availability of extensive genealogic information enables one to sample only individuals descended from the founders of the Central Valley. The genealogical tracing performed to date indicates that the ancestry of most Central Valley BP-I subjects can be delineated at least back to the mid-19th century, and that the vast majority of those sampled will have primarily Central Valley ancestors. Patients who are found to have one or more non-Central Valley ancestors may still be included in a genome search for shared chromosomal segments; the genotype information from a given individual can be weighted depending on the proportion of his or her ancestors who were born in the Central Valley. As migration within the country has steadily increased in the last few decades, it is clear that in the future it will become more difficult to identify individuals whose ancestors were descended predominantly from the Central Valley founders, and larger samples will be required for using LD methods to map disease genes.

The problem presented by individuals who are phenocopies, and whose inclusion would decrease the power to detect genetic linkage, is more difficult to overcome simply by identifying a genetically homogenous study population. In pedigree studies of psychiatric disorders, it is inevitable that a proportion of individuals will be designated with a given phenotype based largely on information elicited in direct structured interviews; for some individuals, who have never or rarely sought medical attention, there may be considerable uncertainty as to whether the presentation and course of illness are "typical" or "classical" for the given syndrome. It is thus an important advantage of LD methods, which draw from the population as a whole, that one can limit the study sample to patients who have well-documented disease phenotypes. For example, among the BP-I individuals whom we have investigated, several clinical variables are typical for the syndrome, as reported in clinical and epidemiological populations that have been studied in North America and Western Europe. As in those samples, we detect an equal prevalence of BP-I in males and females. The distribution in age of onset is quite close to that reported in clinical populations in North America [Baron et al., 1983; Weissman et al., 1988], and includes a secondary peak of later onset (30's and early 40's) in women as noted in clinical populations in Europe [Angst, 1978]. The average length of hospitalization in the sample (M = 16.7 days, F = 17.6 days) is within the range of length of stay for BP patients as reported elsewhere [Verdoux and Bourgeois, 1993], and significantly less than that reported for schizophrenic patients [Sobel et al., 1993]. In summary, for several measurable categories (age of onset, age at first hospitalization, length of stay, and recurrence of episodes), these BP patients were very similar to BP patients studied in North America and Europe [Goodwin and Jamison, 1990].

In comparison to classic LOD score methods, LD mapping through a search for shared segments has several additional advantages when analyzing data for complex disorders such as BP: 1) As with other types of association studies, LD approaches are not nearly as sensitive to misdiagnosis as are LOD score analyses, and do not require knowledge of the parameters of genetic transmission, such as frequency of the disease gene. 2) Although methods have been developed for performing LOD score calculations for two or more causative loci, these require extremely lengthy computation times and have not yet been successfully employed in genome screening studies of humans. In contrast, the methods used to detect shared segments do not differ whether one or several causative loci are present in a population. 3) When shared chromosomal segments are found by LD mapping, one can continue to narrow the candidate region by increasing the sample of patients from a population, and identifying progressively smaller shared segments. By contrast, for a given set of pedigrees, LOD score analysis is useless once a region with no recombinations has been identified; further narrowing requires collection of complete additional families to search for recombinants.

The recent demonstration that a search for shared segments can be used to map disease genes reveals the practical advantage that LD analysis has over other types of association studies. Because the approach is based on the identification of shared haplotypes in patients, it is not necessary to compare the frequencies of particular alleles in patients and normal individuals. As long as one chooses markers that are reasonably polymorphic (i.e., with a polymorphism information content [PIC] >.90), it is very unlikely that a shared haplotype will occur by chance in affected individuals [Houwen et al., 1994]. Moreover, even in the case of low polymorphic content, candidate regions will still be located by LD analysis, though at a lesser specificity (i.e., the sensitivity of the test is no lower). It is thus not necessary to evaluate allelic distributions in the background population for each marker examined, in order to locate a candidate region or regions. In contrast to LD studies, single-marker association studies must compare any allele frequency in affecteds to the allelic distribution at each marker site for the population being studied, a difficult and costly task when performing an entire genome search.

Recent linkage findings for BP have not been confirmed by either reevaluation of the same pedigrees with new markers or evaluation of independent samples. Such replication will thus be an important standard of proof for future reports of linkage. Given the likely etiologic heterogeneity for BP, it may be difficult to replicate findings using samples from different populations. LD analyses may play a valuable role in such situations; identification of chromosome segments shared by patients drawn from the same population as families used for LOD score analyses can provide confirmation for linkage results.

Future efforts will include genotyping of the BP-I patients described above, using markers covering the human genome at about 5–10-cM intervals. Identification of chromosomal segments shared by the affected individuals will suggest the location of BP susceptibility loci in the Costa Rican population. Concurrent LOD score analyses of families in Costa Rica are also expected to identify a candidate region or regions for BP linkage. Either method alone may find statistically significant evidence for linkage of BP to a chromosome region or regions. Comparison of the results obtained using both the LOD score and LD methods will highlight the chromosomal regions that are most likely to contain a locus or loci promoting susceptibility to BP.

ACKNOWLEDGMENTS

This research was supported by the American Psychiatric Association's Program for Minority Research Training in Psychiatry (M.E.), the Wolfe Fund of the Langley Porter Psychiatric Institute (M.E.), the National Alliance for Research in Schizophrenia and Depression Young Investigator's Award (N.F.), and National Institutes of Health grants MH 00916 and MH 49499 (N.F.). We also acknowledge Carmen Araya Ortiz, who helped with the collation and recording of data, and Sue Service, who helped design the

figures and tables. Alison McInnes and Laura Bull provided helpful suggestions during the writing of this paper.

REFERENCES

Abarracin-Gonzalez P (1978): Estudios historico-demografico; algunos familias de la valle central 1650–1850. San Jose, Costa Rica: Research Thesis, Universidad de Costa Rica.

Acosta Vega AC (1980): Breve resena arqueologica y etnologica de Costa Rica, Ministerio de cultura, Juventid y deportes Departamento de Anthropologia, pp 1–45.

Angst J (1978): The course of affective disorders II. Typology of bipolar manic-depressive illness. Arch Psychiatr Nervenkrankh 226: 65–73.

Asociacion Demografica Costaricense (1985): "Demografica y Epidemiologia en Costa Rica." San Jose, Costa Rica: Asociacion Demografica Costaricense, pp 1–30.

Baron M, Risch N, Mendlewicz J (1983): Age at onset in bipolar-related major affective illness: Clinical and genetic implications. J Psychiatr Res 17:5–18.

Beckmann JS, Richard I, Hillaire D, Broux O, Antignac C, Cann H, Cottingham RW Jr, Feingold N, Feingold J, Kalil J, Lathrop M, Marcadet A, Masset M, Mignard C, Passos-Bueno MR, Pellerain N, Zatz M, Dausset J, Fardeau M, Cohen D (1991): A gene for limb-girdle muscular dystrophy maps to chromosome 15 by linkage. C R Acad Sci [III] 312:141–148.

Bertelsen A, Narvald B, Nauge M (1977): A Danish twin study of manic-depressive disorders. Br J Psychiarty 130:330–351.

Cannon-Albert LA, Goldgar DE, Meyer LJ, Lewis Cm, Anderson DE, Fountain JW, Hegi ME, Wiseman RW, Petty EM, Vale AE, et al. (1992): Assignment of a locus for familial melanoma, MLM, to chromosome 9p13-p22. Science 258:1148–1152.

Davies JL, Kawaguchi Y, Bennett ST, Copeman JB, Cordell HJ, Pritchard LE, Reed PW, Gough SCL, Jenkins SC, Palmer SM, Balfour KM, Rowe BR, Farrall M, Barnett AH, Bain SC, Todd JA (1994): A genome-wide search for human type 1 diabetes susceptibility genes. Nature 371:130–136.

De la Chapelle A (1993): Disease gene mapping in isolated human populations: The example of Finland. J Med Genet 30:857–865.

Endicott J, Spitzer RL (1979): Use of the Research Diagnostic Criteria and the Schedule for Affective Disorders and Schizophrenia to study affective disorders. Am Psychiatry 136(1):52–56.

Fernandez ME, Schmidt A, Bassauri V (1976): La poblacion en Costa Rica, Editorial Universidad de Costa Rica. San Jose, Costa Rica: Ciudad Universitaria Rodrigo Facio: Universidad de Costa Rica, Instituto de Investigaciones Sociales, pp 5–103.

Fourgerousse F, Broux O, Richard I, Allamand V, Pereira de Souza A, Bourg N, Brenguier L, Devaud C, Pasturaud P, Roudat C, Chiannilkulchai N, Hillaire D, Bui H, Chumakov I, Weissenbach J, Cherif D, Cohen D, Beckman J (1994): Mapping of a chromosome 15 region involved in limb girdle muscular dystrophy. Hum Mol Genet 3:285–293.

Goodwin FK, Jamison KR (1990): The manic-depressive spectrum. In Goodwin FK, Jamison KR (eds): "Manic-Depressive Illness." New York: Oxford University Press, pp 77–84.

Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, Bernardi G, Lathrop M, Weissenbach J (1994): The 1993–94 Genethon human genetic linkage map. Nat Genet 7:246–339.

Heyer E (1994): Genetic consequences of differential demographic behavior in the Saguenay region. Am J Hum Genet [Suppl] 55:153.

Houwen RHJ, Baharloo S, Blankenship K, Raeymakers P, Juyn J, Sandkuijl LA, Freimer NB (1994): Genome screening by searching for shared segments: Mapping a gene for benign recurrent intrahepatic cholestasis. Nat Genet 8:380–386.

Kramer M, Sartorius N, Jablensky A, Gulbinat W (1979): The ICD-9 classification of mental disorders: A review of its development and contents. Acta Psychiatr Scand 59:241–262.

Lander E, Botstein D (1986): Mapping complex genetic traits in humans: New methods using a complete RFLP linkage map. Cold Spring Harbor Symp Quant Biol LI:49.

Leon PE, Raventos H, Lynch E, Morrow J, King MC (1992): The gene for an inherited form of deafness maps to chromosome 5q31. Proc Nat Acad Sci USA 89:5181–5184.

Lowell Gudmundson Kristjanson MA (1978): "Estratificacion Socio-Racial y Economica de Costa Rica: 1700–1850." San Jose, Costa Rica: Editorial Universidad Estatal a Distancia pp 90–125.

Mathieu J, De Braekeleer M, Prevost C (1990): Geneological reconstruction of myotonic dystrophy in the Saguenay-Lac-Saint-Jean area (Quebec, Canada). Neurology 40:839–842.

Melendez C (1982): "Conquistadores y Pobladores: Origenes Historico-Sociales de los Costarricenses." San Jose, Costa Rica; Editorial Universidad Estatal a Distancia, pp 1–133.

Mendlewicz J (1976): The age factor in depressive illness. J Gerontol 31:300–303.

Molina Jimenez, I (1991): Costa Rica 1800–1850: El legado colonial y la genesis del capitalismo. Editorial de la Universidad de Costa Rica. San Jose, Costa Rica, pp 60–88.

Nurnberger JI, Blehar MC, Kaufman CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, Severe JB, Malaspina D, Reich T (1994): Diagnostic interview for genetic studies: Rationale, unique features and training. Arch Gen Psychiatry 51:849–859.

Nystrom-Lahti M (1994): Close linkage to chromosome 3p and conservation of ancestral founding haplotype in hereditary nonpolyposis colorectal cancer families. Proc Natl Acad Sci USA 91:6034–6058.

Pauls DL (1993): Behavioral disorders: Lessons in linkage. Nat Genet 3:4–5.

Pauls DL, Morton LA, Egeland JA (1992): Risks of affective illness among first-degree relatives of bipolar I old-order Amish probands. Arch Gen Psychiatry 49(9):703–708.

Peltomaki P, Raltonen LA, Sistonen P, Pylkkanen L, Mecklin JP, Jarvinen H, Green JS, Jass JR, Weber JL, Leach FS (1993): Genetic mapping of a locus predisposing to human colorectal cancer. Science 260:810–812.

Saborio M (1992): Experience in Providing Genetic Services in Costa Rica. New York: Alan R. Liss, Inc., for the National Foundation—March of Dimes. BD:OAS XXVIII (3):96–102.

Sanabria VM (1977): Geneologias de Cartago Hasta 1850. In Tinoco LD (ed): "La Poblacion de Costa Rica y Origenes de los Costaricenses." San Jose, Costa Rica: Editorial Costa Rica. 1977, pp 15–214.

Smith R, Rug W, Tobriner G, Won W-J (1978): Linkage analysis. Hum Genet 17:2–35.

Sobel JL, Heston LL, Sommer S (19930: Novel association approach for determining the genetic predisposition to schizophrenia: Case-control resource and testing of a candidate gene. Am J Med Genet 48:28–35.

Spence MA, Ameli H, Sadovnick AD (1993): A single major locus is the best explanation for bipolar family data: Results of complex segregation analysis. Am J Hum Genet 53:862.

Stone K (1989): Mania in the elderly. Br J Psychiatry 155:220–224.

Te Meerman G, Van der Meulen M, Sandkuijl, L (1994): Expected size of shared halotypes surrounding a common disease gene. Am J Hum Genet 55(3)suppl:A205

Thiel BA (1951): "Monografia de la Poblacion de la Republica de Costa Rica en el Siglo XIX." San Jose, Costa Rica: Ministerio de Economia y Hacienda, Direccion General de Estadistica y Censos, pp 1–43.

Tsuang M, Faraone S, Green A (1994): Genetic epidemiology of mood disorders. In Papolos DF, Lachman HM (eds): "Genetic epidemiology of mood disorders. In Papolos DF, Lachman HM (eds): "Genetic Studies in Affective Disorders." New York: John Wiley and Sons, Inc., pp 1–75.

Uhrhammer NA, Lange E, Gatti RA (1994): Localization of an ataxia-telangiectasia gene by linkage disequilibrium in the Costa Rican population. Am J Hum Genet [Suppl] 55:355.

Verdoux H, Bourgeois M (1993): A comparison of manic patient subgroups. J Affective Disord 27:267–272.

Weissman MM, Leaf PJ, Tischler GL, Blazer DG, Karno M, Bruce ML, Florio LP (1988): Affective disorders in five United States communities. Psychol Med 18:141–153.